Claims

What is claimed is:

1. A method of automatically processing an input sequence of data symbols, the method comprising the steps of:

identifying at least one regularly identifiable expression in the input sequence of data symbols;

identifying at least a portion of information associated with the at least one regularly identifiable expression; and

extracting the portion of information.

2. The method of claim 1, wherein the regularly identifiable expression identifying step comprises comparing the input sequence of data symbols to one or more previously-stored regularly identifiable expressions to determine if there is a match between a portion of the input sequence and at least one of the previously-stored regularly identifiable expressions.

3. The method of claim 1, further comprising the step of normalizing the input sequence of data symbols prior to identifying the regularly identifiable expression.

4. The method of claim 1, further comprising the step of identifying one or more classes of data symbols in the input sequence prior to identifying the regularly identifiable expression.

5. The method of claim 1, wherein the at least one regularly identifiable expression comprises a characteristic phrase that typically precedes a particular portion of information.

6. The method of claim 1, wherein the at least one regularly identifiable expression comprises a characteristic phrase that typically follows a particular portion of information.

7. The method of claim 1, wherein the extracted portion of information is used to take a specified action.

8. The method of claim 1, wherein the extracted portion of information is at least one of visually and audibly presented to the user.

9. The method of claim 1, wherein the regularly identifiable expression identifying step is performed in accordance with one or more programs written in one of the flex, lex, and perl programming language.

10. The method of claim 1, wherein the input sequence of data symbols is representative of at least one of text data, transcribed spoken data, deoxyribonucleic acid sequence data, ribonucleic acid sequence data, amino-acid sequence data, audio sequence data, and video sequence data.

11. The method of claim 1, wherein the input sequence of data symbols is representative of a voice mail message.

12. Apparatus for automatically processing an input sequence of data symbols, the apparatus comprising:

at least one processor operative to: (i) identify at least one regularly identifiable expression in the input sequence of data symbols; (ii) identify at least a portion of information associated with the at least one regularly identifiable expression; and (iii) extract the portion of information; and

memory, operatively coupled to the at least one processor, for storing at least a portion of results associated with the identifying and extracting operations.

13. The apparatus of claim 12, wherein the regularly identifiable expression identifying operation comprises comparing the input sequence of data symbols to one or more regularly identifiable expressions, previously stored in the memory, to determine if there is a match between a portion of the input sequence and at least one of the previously-stored regularly identifiable expressions.

14. The apparatus of claim 12, wherein the at least one processor is further operative to normalize the input sequence of data symbols prior to identifying the regularly identifiable expression.

15. The apparatus of claim 12, wherein the at least one processor is further operative to identify one or more classes of data symbols in the input sequence prior to identifying the regularly identifiable expression.

16. The apparatus of claim 12, wherein the at least one regularly identifiable expression comprises a characteristic phrase that typically precedes a particular portion of information.

17. The apparatus of claim 12, wherein the at least one regularly identifiable expression comprises a characteristic phrase that typically follows a particular portion of information.

18. The apparatus of claim 12, wherein the extracted portion of information is used to take a specified action.

19. The apparatus of claim 12, wherein the extracted portion of information is at least one of visually and audibly presented to the user.

20. The apparatus of claim 12, wherein the regularly identifiable expression identifying operation is performed in accordance with one or more programs written in one of the flex, lex, and perl programming language.

21. The apparatus of claim 12, wherein the input sequence of data symbols is representative of at least one of text data, transcribed spoken data, deoxyribonucleic acid sequence data, ribonucleic acid sequence data, amino-acid sequence data, audio sequence data, and video sequence data.

22. The apparatus of claim 12, wherein the input sequence of data symbols is representative of a voice mail message.

23. An article of manufacture for use in automatically processing an input sequence of data symbols, comprising a machine readable medium containing one or more programs which when executed implement the steps of:

identifying at least one regularly identifiable expression in the input sequence of data symbols;

identifying at least a portion of information associated with the at least one regularly identifiable expression; and

extracting the portion of information.

24. Apparatus for automatically processing an input sequence of data symbols, the apparatus comprising:

a data capture device for obtaining the input sequence of data symbols;

at least one processor, operatively coupled to the data capture device, operative to: (i) identify at least one regularly identifiable expression in the input sequence of data symbols; (ii) identify at least a portion of information associated with the at least one regular expression; and (iii) extract the portion of information;

memory, operatively coupled to the at least one processor, for storing at least a portion of results associated with the identifying and extracting operations; and

a data output device, operatively coupled to the at least one processor, for presenting the extracted portion of information to a user.

25. A method of automatically processing an input document, the method comprising the steps of:

identifying one or more regular expressions in the input document by comparing the input document to one or more previously-stored regular expressions to determine if there is a match between a portion of the input document and at least one of the previously-stored regular expressions; and

identifying at least a portion of information associated with the at least one regular expression for extraction.